SPACE AND THE CHALLENGES TO OCCUPY IT

Lorenzo Dinia

07/27/2024

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Conclusions

EXECUTIVE SUMMARY

SpaceX	Space exploration business is highly competitive. In this race, SpaceX was the first private company to return a spacecraft from low-earth orbit in Dec 2010
	Falcon 9 first stage introduced the concept of reusable launcher by landing successfully
	Other launch vehicles still present reliability issues, e.g. Falcon Heavy
	SpaceX Falcon 9 reusable first stage -> Save money – 62 million vs 165 millions from other providers

EXECUTIVE SUMMARY

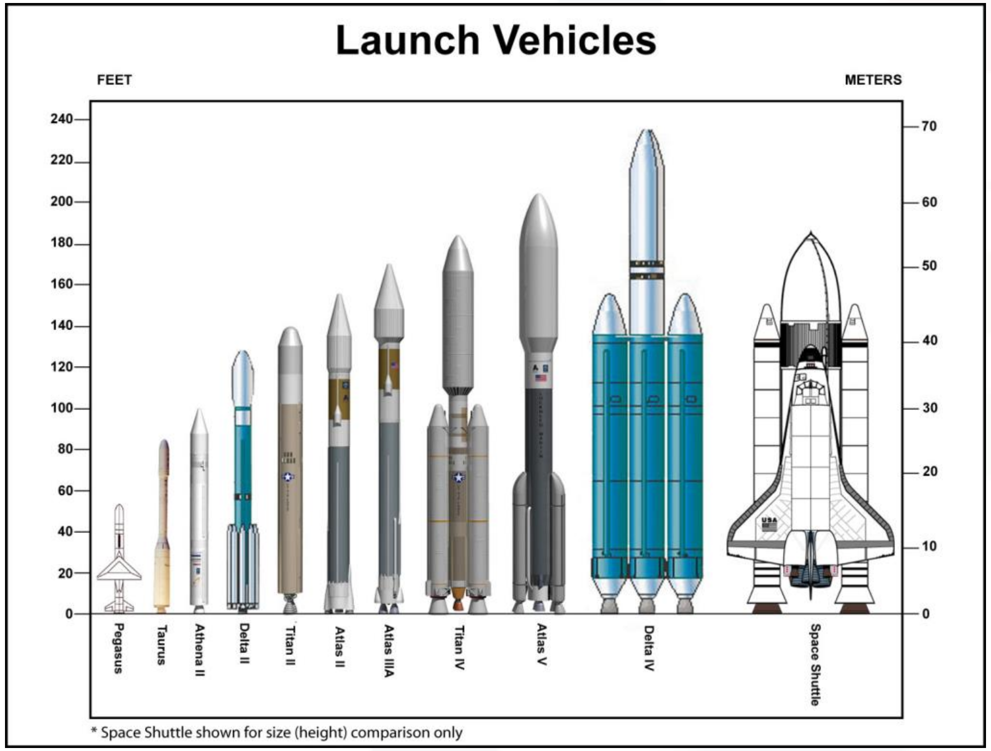
SpaceX

Falcon 9 booster return and recovery depend on factors such as:
- Payload mass
- Booster versions
- Launching sites
- Orbit

The cost of a launch mainly depends on the return of the first stage

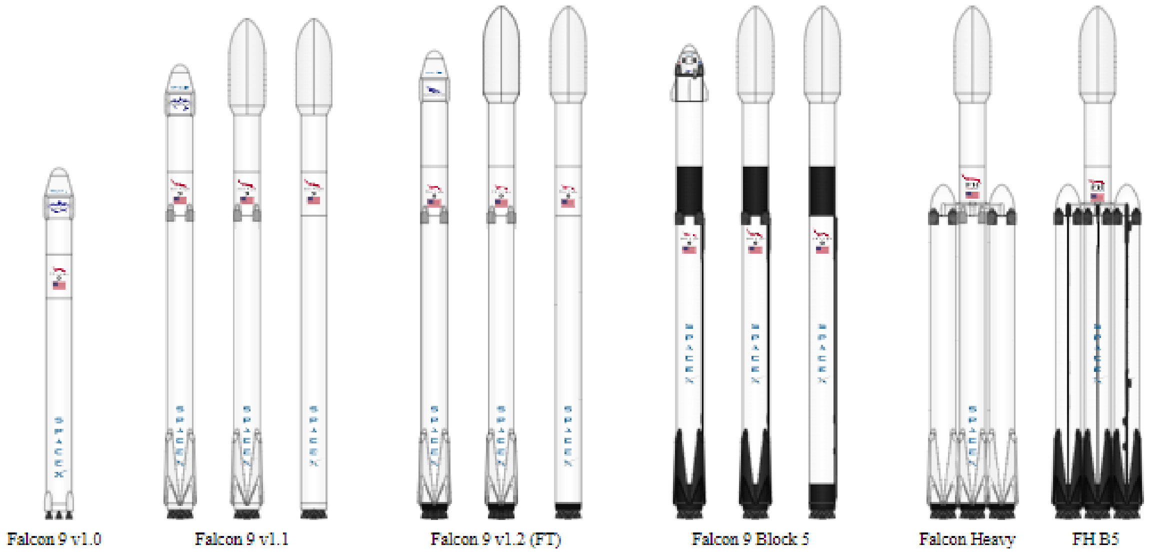
Based on previous launches, it was predicted that Falcon 9 first stage will land successfully with an accuracy of around 84%

INTRODUCTION



- The commercial rocket launches is a very competitive market with the main world economies participating in the race
- The expectation of a continuous economic and population growths is introducing some challenges to humans on a planet characterized by fix boundaries
- In the last few years, the investment related to Space exploration and potentially Space tourism have exponentially increased

INTRODUCTION



- SpaceX Falcon 9 and Falcon Heavy database was used to generate a report in which past data were analyzed and its future success was predicted with the best model

- Collection of Data
  - Request to the SpaceX API
  - Clean the data
- Web scraping is the key to understand Falcon 9 historical launch records
  - Wikipedia is the platform where web scraping is performed
- Exploring and preparing data
  - Exploratory Data Analysis
  - Preparing Data Feature Engineering

METHODOLOGY

Data Collection, Web Scraping

- Request rocket launch data from SpaceX API
- Extract launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

Data Wrangling

- Exploratory Data Analysis
- Determine Training Labels

Data Analysis

- Execute SQL queries
- Data Analysis and Feature Engineering using Panda and Matplotlib

Data Visualization

- Launch sites with Folium
- Success rate with Plotly Dash

Data Prediction

- Standardize the data
- Split the data into training and test data
- Identify the best machine learning prediction method using test data

METHODOLOGY – DATA COLLECTION

Request and Parse Rocket Launch Data from SpaceX API using a GET Request

→ Json Response Content Turned Into a Panda Dataframe

→ Dataframe is filtered to obtain Falcon 9 launches

→ Replace Missing Values with the Mean

→ Dataframe exported to a CSV

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

METHODOLOGY – WEB SCRAPING

- Web Scraping Falcon 9 launch records from Wikipidia
- BeautifulSoup function is used for web scraping
- Used helper functions to process web scraped HTML table
- Scrape the data from a snapshot of the list of Falcon 9 and Falcon heavy launches Wikipage updated on 9th June 2021

Perform an HTTP GET Method to Request the Falcon 9 Launch HTML page as an HTTP Response → Extract all column/variable names from the HTML table header → Create a Dataframe by Parsing the Launch HTML Tables → Dataframe exported to a CSV

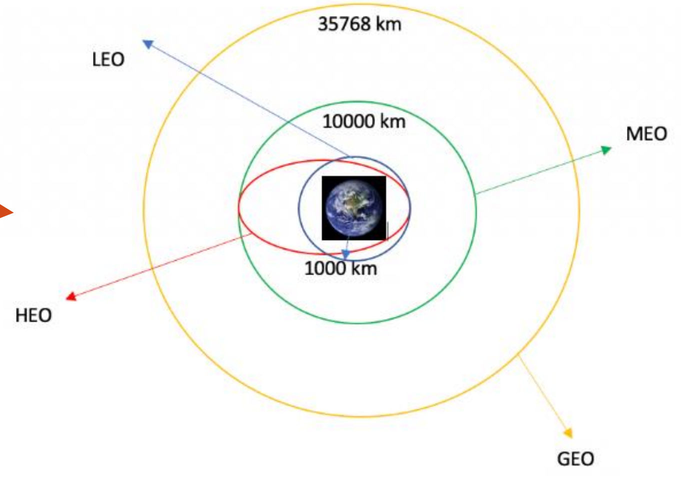
https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-webscraping.ipynb

METHODOLOGY – DATA WRANGLING

- Calculated the number of launches on each site

CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

- Each launch aims to a dedicated orbit – LEO, VLEO, GTO, SSO, ES-L1,HEO, ISS, MEO, HEO, GEO, PO



- Calculated the number and occurrence of each orbit
- Calculated the number and occurrence of mission outcomes of the orbits (True means that the mission outcome was successfully landed, False and None mean that the mission outcome was successfully landed)

True ASDS	41
None None	19
True RTLS	14
False ASDS	6
True Ocean	5
False Ocean	2
None ASDS	2
False RTLS	1

- Created a training label: “Class”

Class = 0 → Booster landing failure

Class = 1 → Booster landing success

ReusedCount	Serial	Longitude	Latitude	Class
0	B0003	-80.577366	28.561857	0
0	B0005	-80.577366	28.561857	0
0	B0007	-80.577366	28.561857	0
0	B1003	-120.610829	34.632093	0
0	B1004	-80.577366	28.561857	0

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-webscraping.ipynb

METHODOLOGY – EDA WITH SQL

SpaceX DataSet are explored by executing SQL queries

- The names of the unique launch sites in the space mission are displayed
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

METHODOLOGY – DATA ANALYSIS

Data analysis is performed via Pandas and Matplotly

- Scatter Plot - Flight Number vs Pay load Mass (kg) with hue → Class
- Scatter Plot - Flight Number vs Launch Site with hue → Class
- Scatter Plot – Pay load Mass (kg) vs Launch Site with hue → Class
- Bar Plot – Orbit vs Class
- Scatter Plot - Flight Number vs Orbit
- Scatter Plot – Pay load Mass (kg) vs Orbit with hue → Class
- Line Plot – Year vs Success Rate

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

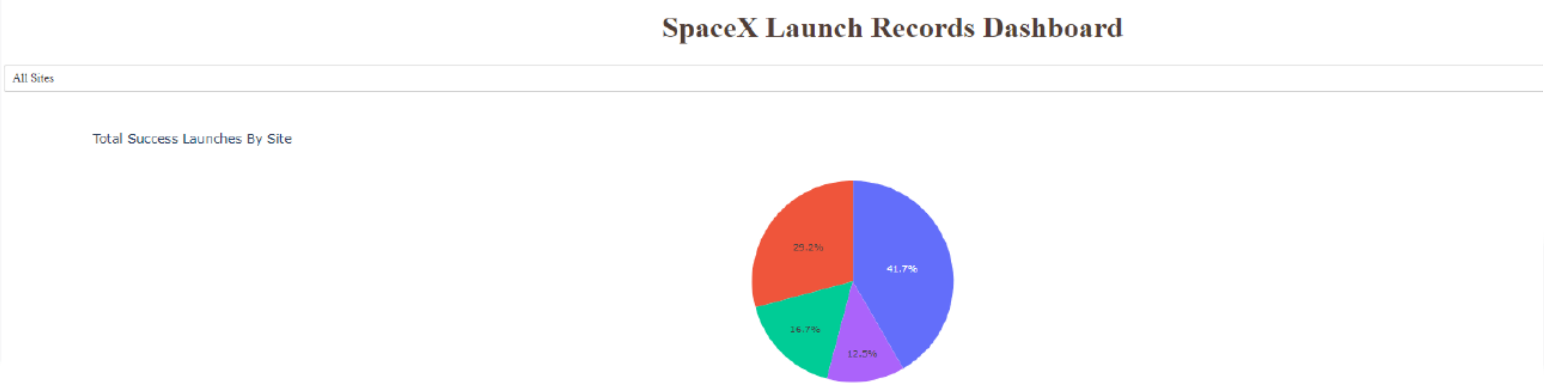
METHODOLOGY – INTERACTIVE MAP WITH FOLIUM

- The launch success rate may depend on many factors such as payload mass, orbit type, booster version. It may also depend on the location and proximities of a launch site
- An interactive map was created with Folium to explore potential relationships between launch success rate and site location
  - All site were marked on the map
  - The success/failed launches for each site were marked
  - The distance between launch sites to their proximities was calculated

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/lab_jupyter_launch_site_location.ipynb

METHODOLOGY – DASHBOARD WITH PLOTLY DASH

- A Dashboard was created to perform interactive visual analytics on SpaceX launch data in real-time
- The following 4 steps are followed to create the Dashboard:
  - Added a Launch Site Drop-down Input Component
  - Added a callback function to render success-pie-chart based on selected site dropdown
  - Added a Range Slider to Select Payload
  - Added a callback function to render the success-payload-scatter-chart scatter plot



https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/spacex_dash_app.py

METHODOLOGY – MACHINE LEARNING PREDICTION

- Perform exploratory Data Analysis and determine Training Labels
  - Create a column for the class
  - Standardize the data
    - Standardize the data in X then reassign it to the variable X using *transform = preprocessing.StandardScaler()*
  - Split into training data and test data
    - Use the function train_test_split to split the data X and Y into training and test data

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2,random_state=2)
print ('Train set:', X_train.shape, Y_train.shape)
print ('Test set:', X_test.shape, Y_test.shape)
Train set: (72, 83) (72,)
Test set: (18, 83) (18,)

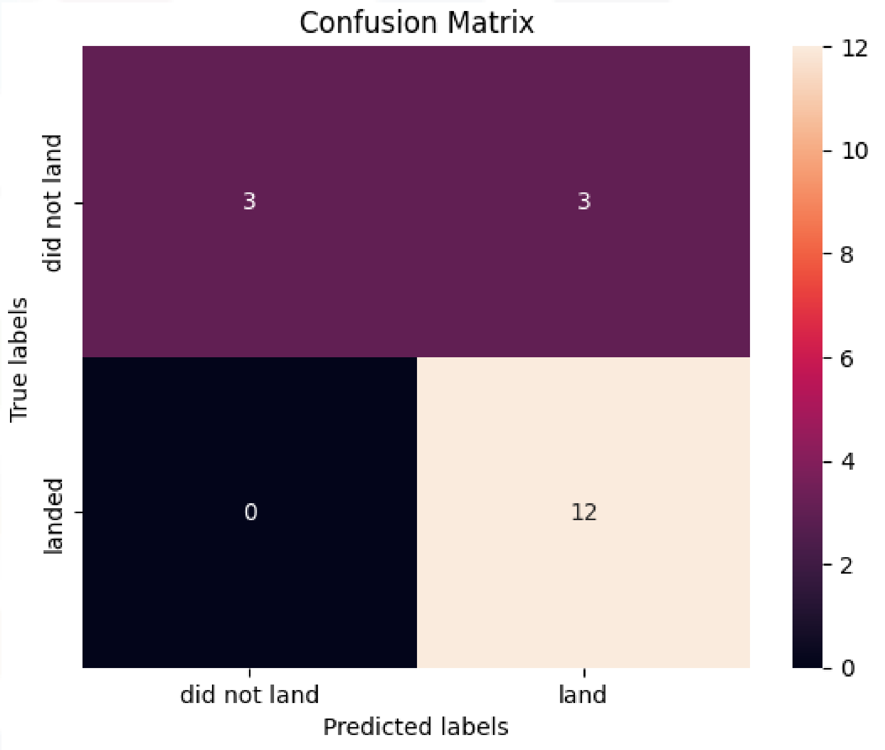
- Find the best Hyperparameter for SVM (Support Vector Machine), Classification Trees, KNN (K Nearest Neighbors) and Logistic Regression
  - Identify the best method for the prediction by using test data

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

METHODOLOGY – MACHINE LEARNING PREDICTION

Predict analysis → Accuracy confusion matrix

Selection of the best model



https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

RESULTS

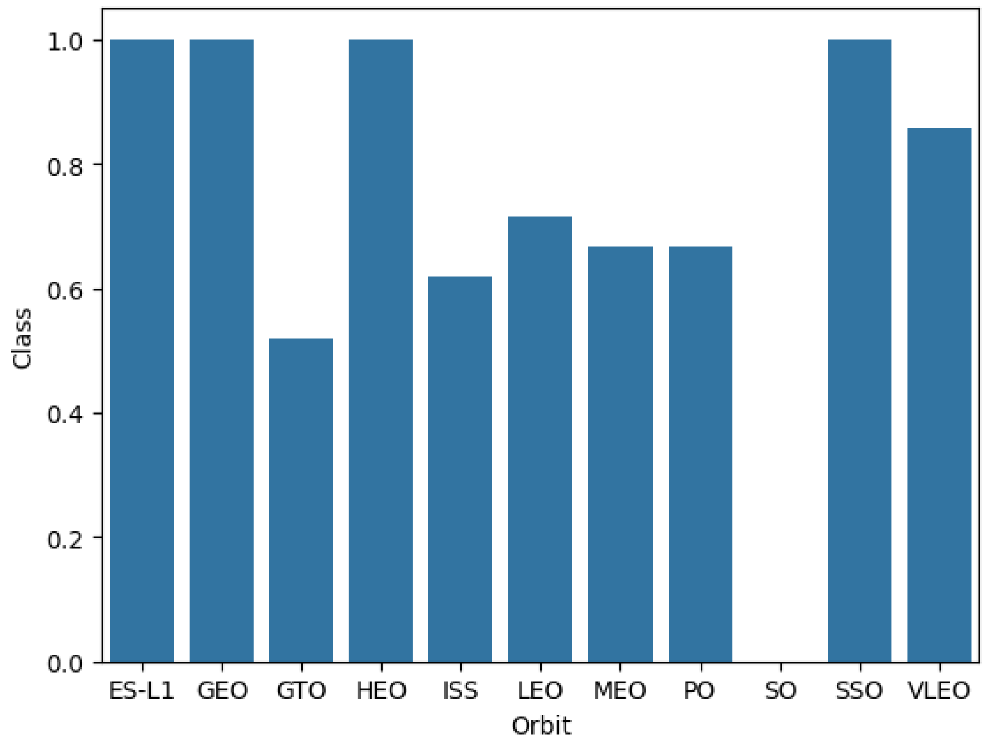
- Exploratory data analysis results
- SQL queries results
- Interactive map with Folium results
- Dashboard results
- Predictive analysis results

RESULTS

The number of launches per orbit and the success rate per orbit are displayed

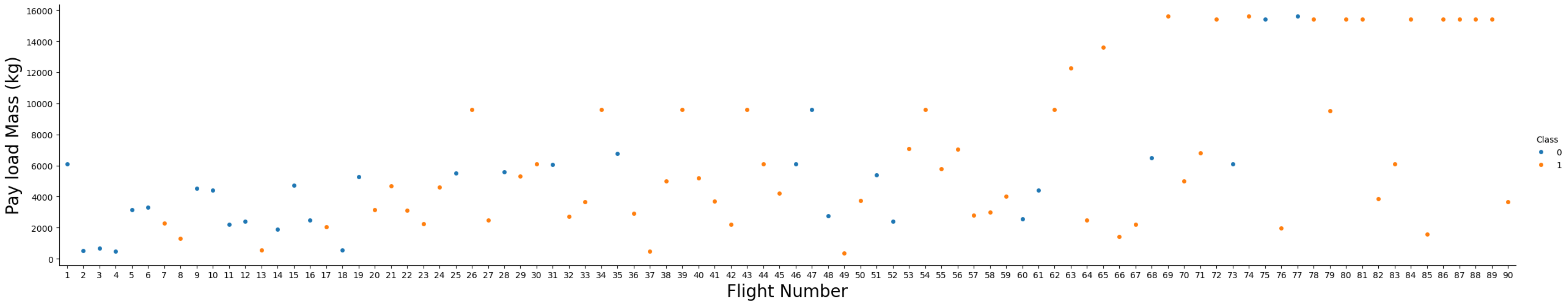
- Number of launches per orbit
  - GTO 27
  - ISS 21
  - VLEO 14
  - PO 9
  - LEO 7
  - SSO 5
  - MEO 3
  - ES-L1 1
  - HEO 1
  - SO 1
  - GEO 1

Success rate per orbit



RESULTS – EDA

Flight Number vs Pay load Mass (kg)

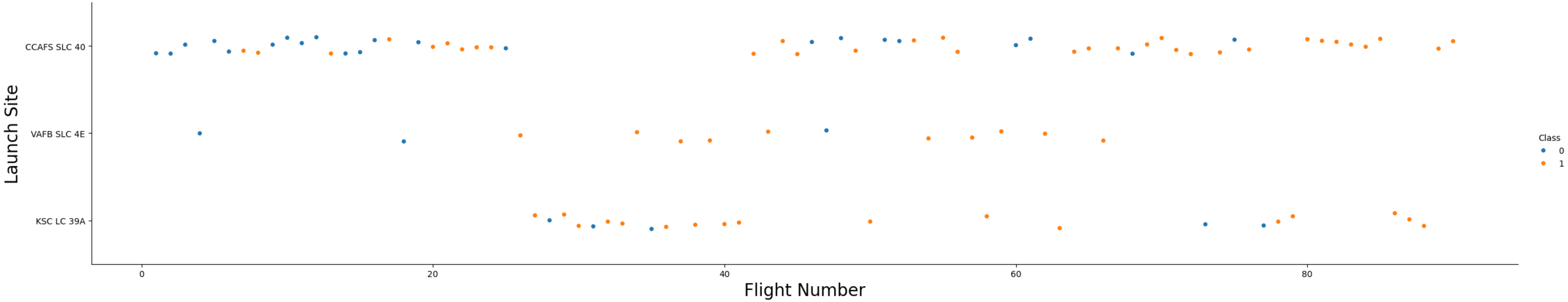


- Increasing the flight number, there is an increase of success rate with low pay load mass
- Only for an high flight number was used a pay load mass around 16,000 kg. Out of 13 launches there were 11 success and 2 failures
- Low pay load masses were used for early stage of the Falcon9 project
- The more mature the Falcon 9 project became, the more heterogeneity in the pay load mass used is displayed

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/edadataviz.ipynb

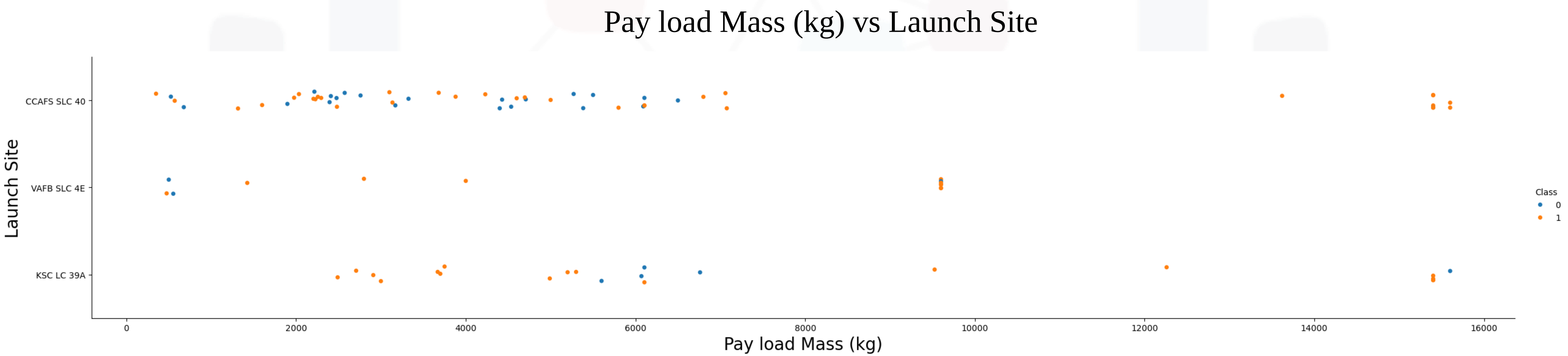
RESULTS – EDA

Flight Number vs Launch Site



- Increasing the flight number, there is an increase of success rate for launch site VAFB SLC 4E
- Launch site VAFB SLC 4E was used exclusively for 13 launches, a small number compared the number of flights at CCAFS SLC 40
- The success rate increases for each launch site proportionally to the number of flights

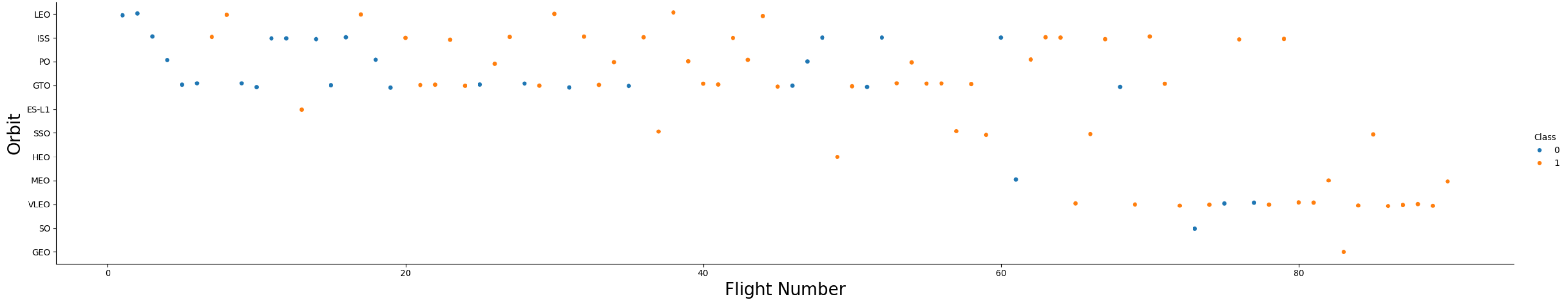
RESULTS – EDA



- The lower the pay load mass, the higher the success rate of the launch site KSC LC 39A is
- Higher is the pay load mass, higher is the success rate of the launch site CCAFS SLC 40. All launches with pay load mass around 16,000 Kg had the first stage successfully returning

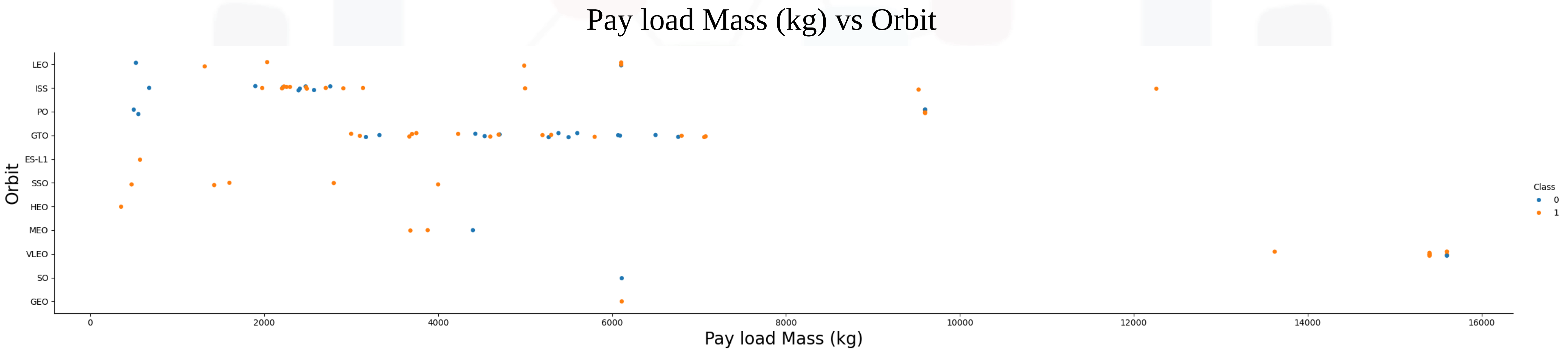
RESULTS – EDA

Flight Number vs Orbit



- The success rate is higher for higher flight number for LEO
- All flights for SSO orbit were successful
- There is only one flight related to GEO orbit and was successful
- There is only one flight related to SO orbit and failed

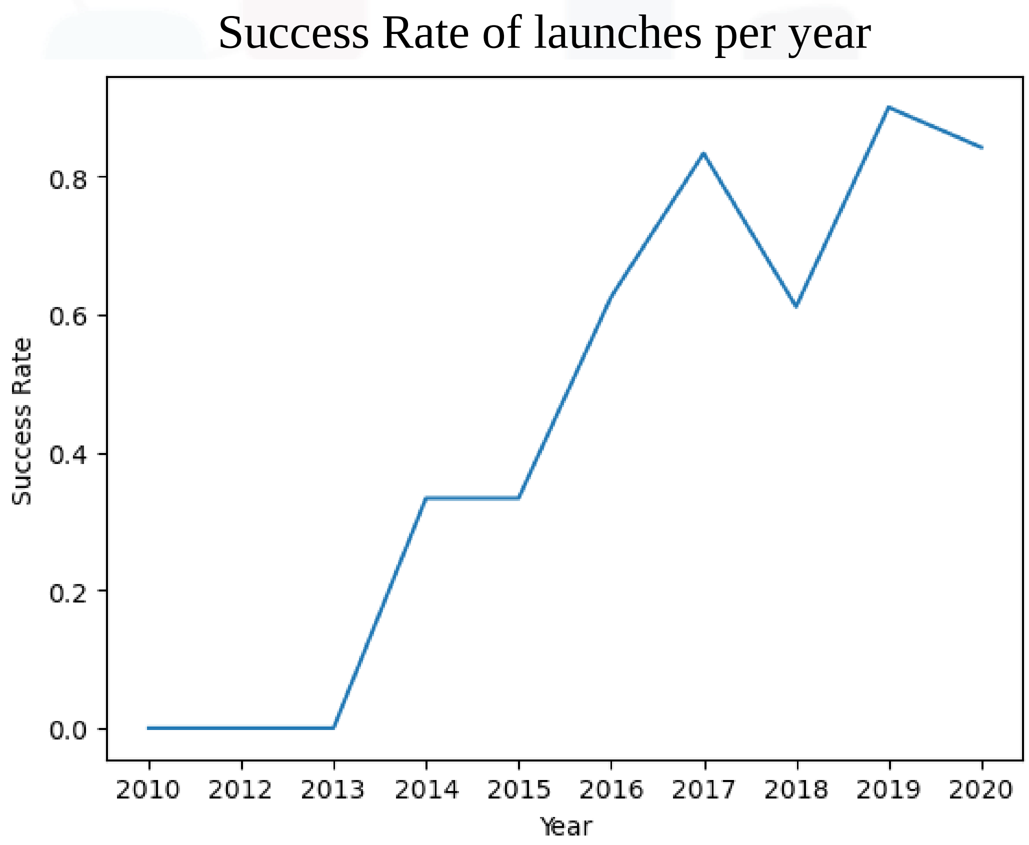
RESULTS – EDA



- Po, LEO and ISS: there is higher success rate for higher pay load mass (kg)
- All flights in SSO were successfully independently by the pay load mass
- There is one flight for ES-L1 and HEO orbits, both were successful

RESULTS – EDA

- Considering the total number of launches the average success rate is 0.6667
- The success rate gradually increased since 2013 until 2020



RESULTS – EDA WITH SQL

- Displayed the names of the unique launch sites in the space mission
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

RESULTS – EDA WITH SQL

- Listed the total number of successful and failure mission outcomes
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- The first successful landing outcome in ground pad was achieved on 12/22/2015
- The total payload mass carried by boosters launched by NASA was 619,967 Kg
- The average payload mass carried by boosted version F9 v1.1 was 2,928.4

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Landing_Outcome	Count_of_Landing
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

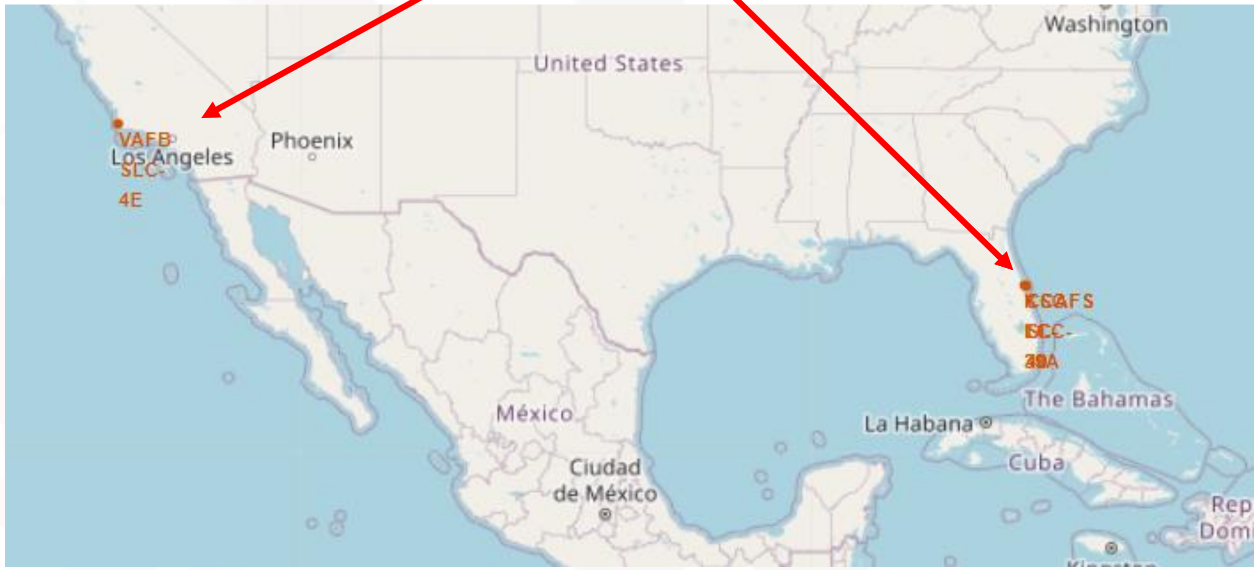
https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

RESULTS – LAUNCH SITES

The Launch sites for SpaceX were 4:

- CCAFS LC-40: Cape Canaveral Launch Complex 40 (FL)
- CCAFS SLC-40: Cape Canaveral Space Launch Complex 40 (FL)
- KSC LC-39A: Kennedy Space Center – Merritt Island (FL)
- VAFB SLC-4E: Vandenberg Space Launch Complex 4 (CA)

Launch Sites



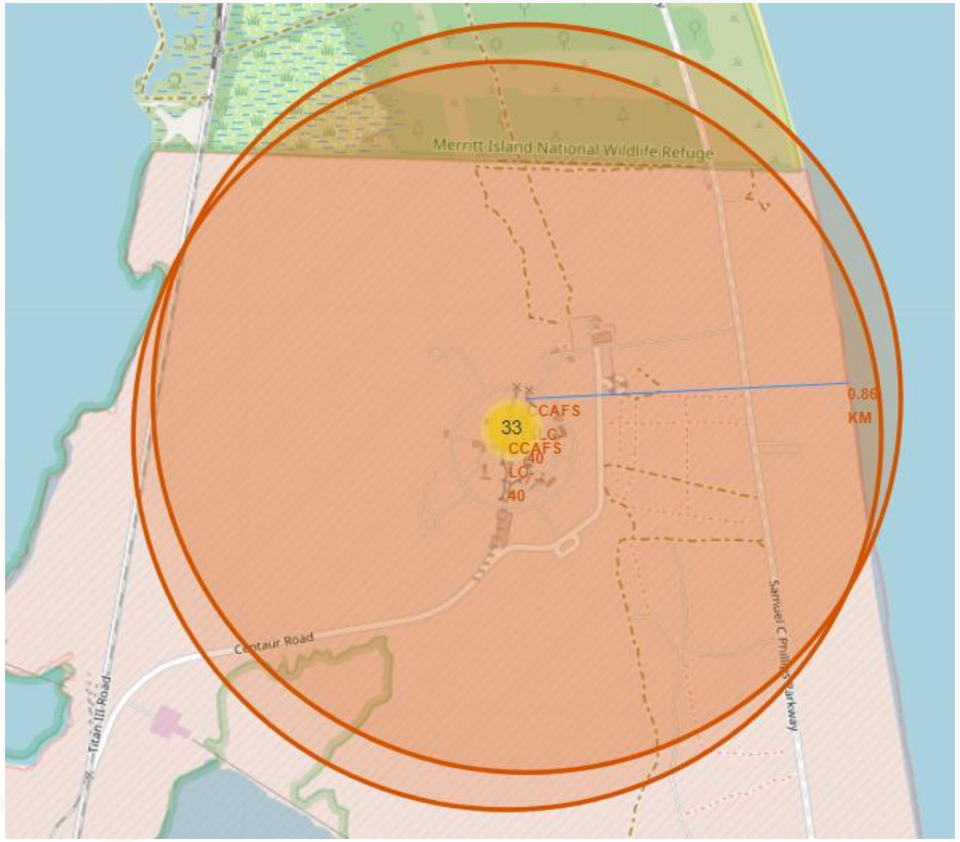
	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/lab_jupyter_launch_site_location.ipynb

RESULTS - LAUNCH SITES

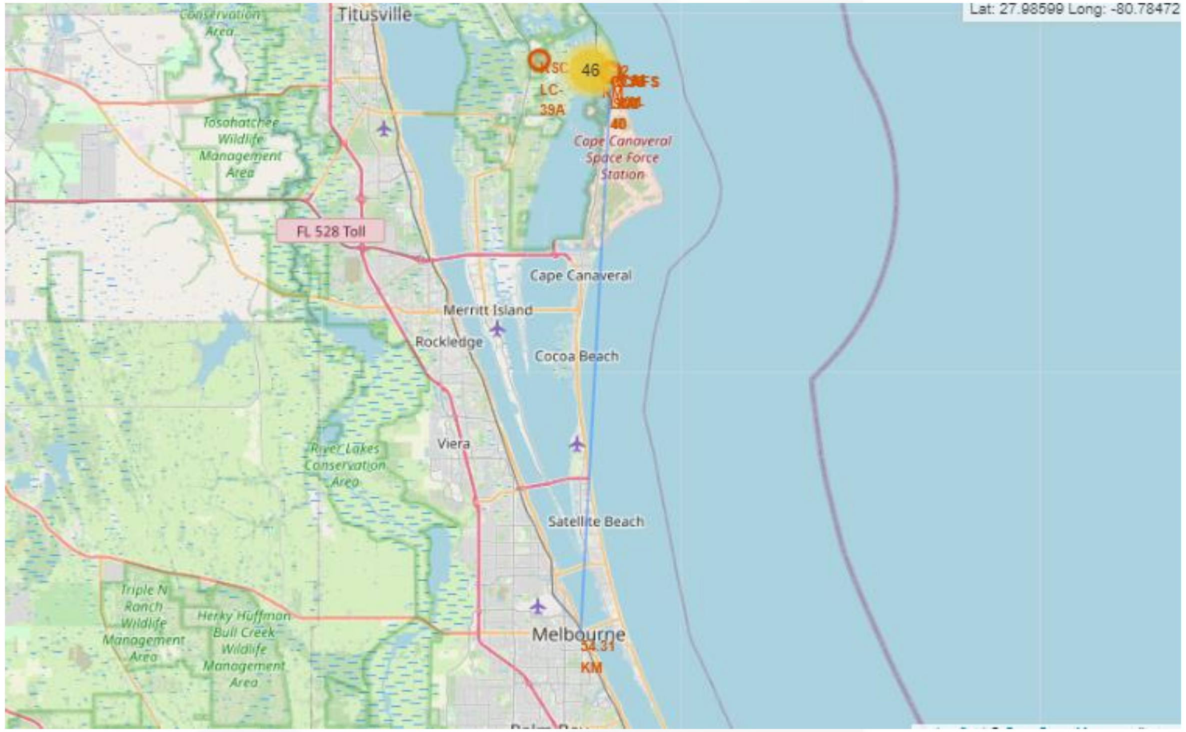
- Launch Site - CCAFS SLC 40 – is located in Florida and it is only 0.862 Km far from the coastline

Launch Site - CCAFS SLC 40



RESULTS - LAUNCH SITES

- The distance between the launch site CCAFS SLC-40 and the city Melbourne is 54.31 Km
- The distance between the launch site CCAFS SLC-40 and the Nasa Railroad is 1.22 Km
- The distance between the launch site CCAFS SLC-40 and the Samuel Phillips Parkway is 0.587 Km



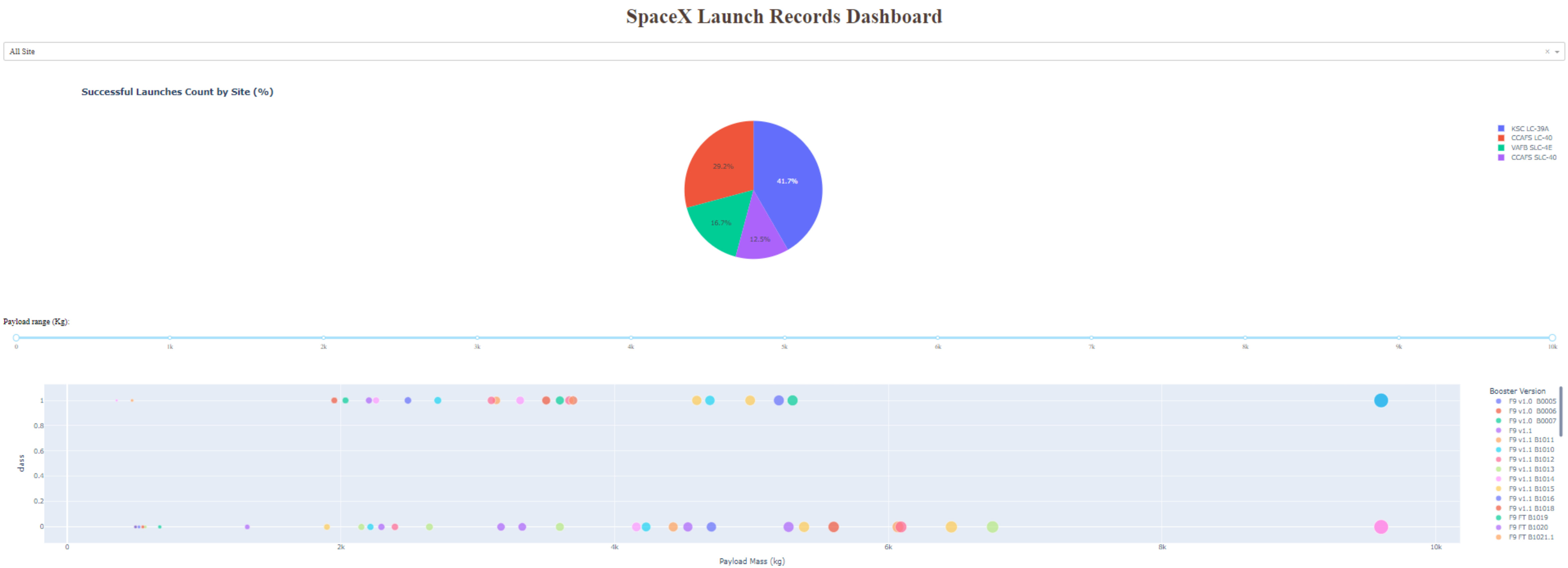
RESULTS - LAUNCH SITES

The launch site is a factor in determining the success rate

- They are close to railways and highways. Most likely, this is due to transport heavy parts of the rockets
- There is also a certain distance from cities and this is due to safety reasons
- Launch sites are close to the coast because in case of loss of control the rocket can be let pointing the water and also it minimizes the risk of having any debris dropping or exploding near people
- Launch sites are close to the Equator line because a rocket can take advantage of the Earth’s rotational speed. Near the equator, a rocket is already moving a speed of over 1,650 Km/h relative to the Earth’s center

DASHBOARD

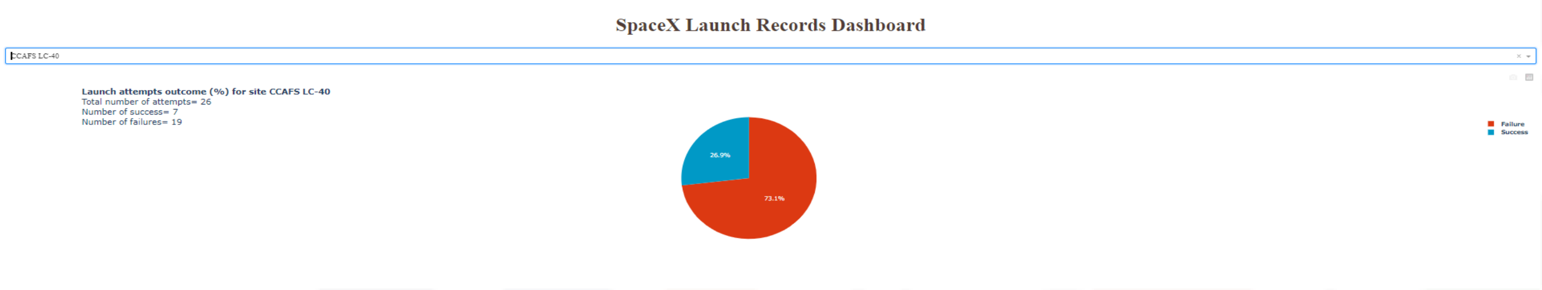
Successful Launches Count by Site (%)



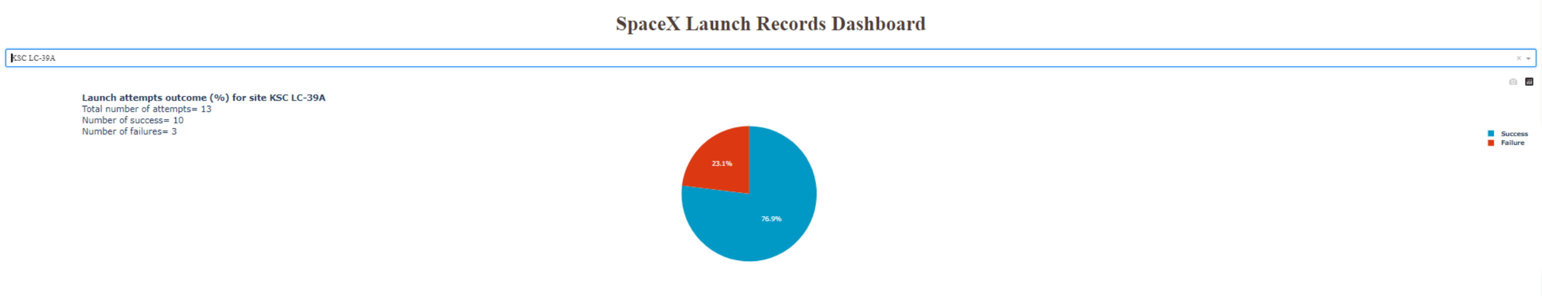
Success and failed missions of different Booster Version per different Payload Mass (Kg)

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/spacex_dash_app.py

DASHBOARD

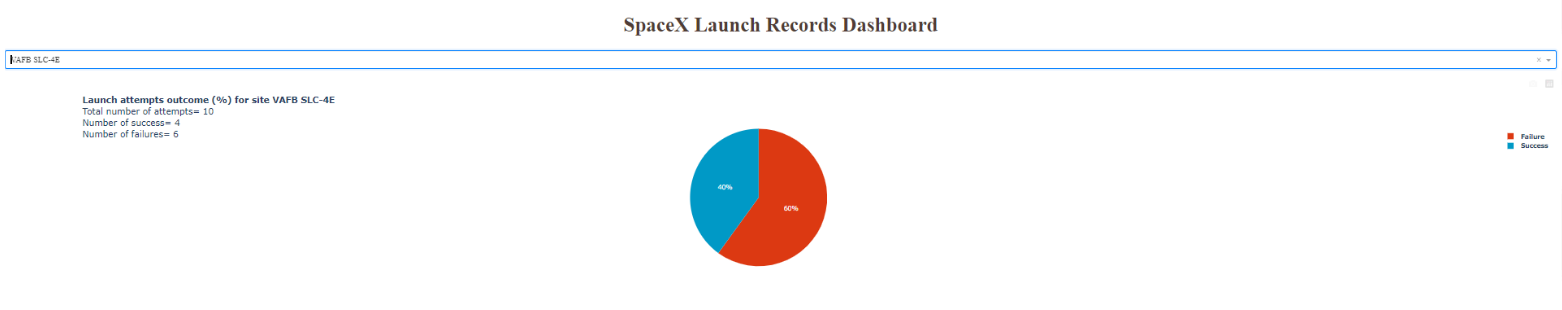


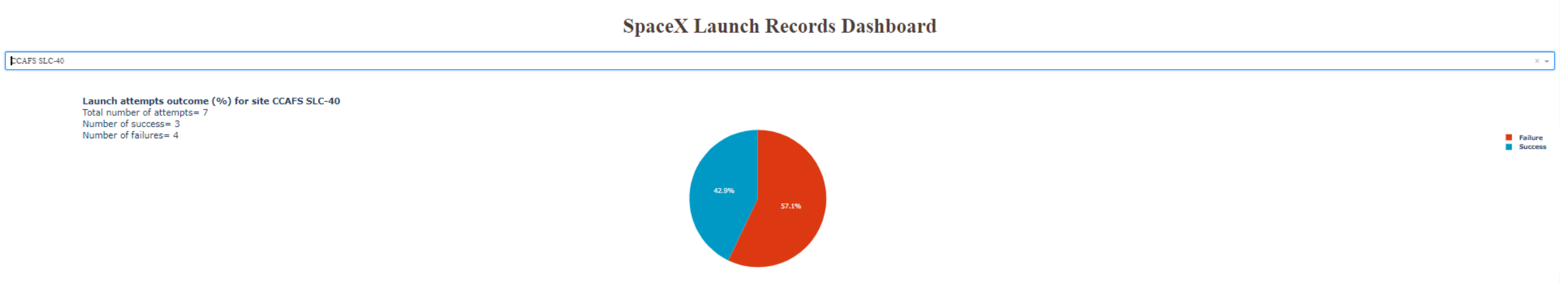
CCAFS LC-40
Number of success = 7
Number of failures = 19



KSC LC-39A
Number of success = 10
Number of failures = 3

DASHBOARD





VAFB SLC-4E
Number of success = 4
Number of failures = 6

CCAFS SLC-40
Number of success = 3
Number of failures = 4

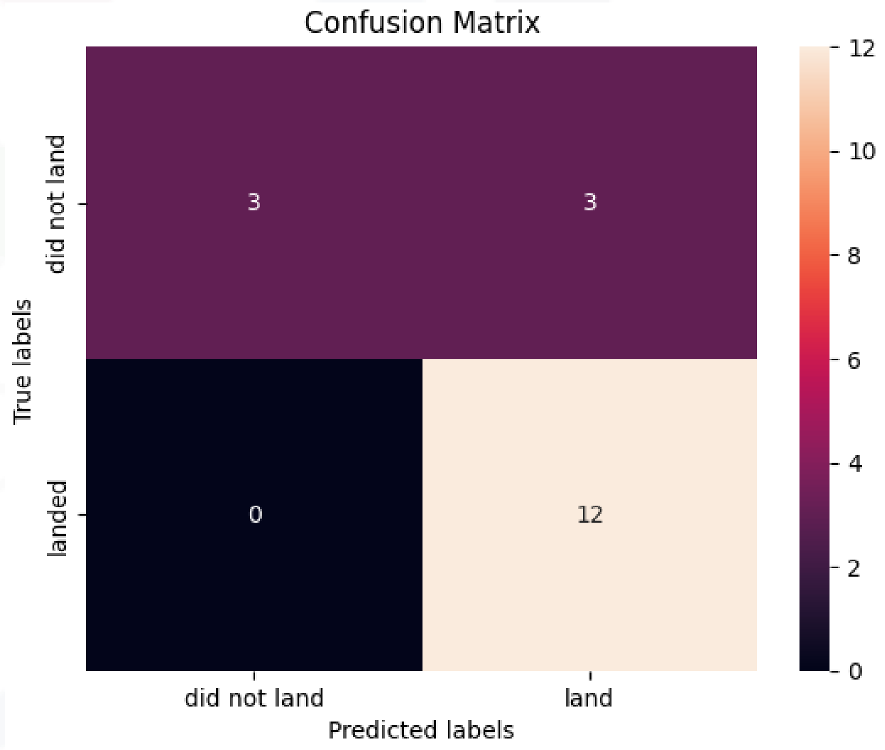
RESULTS – MACHINE LEARNING PREDICTION

- Linear Regression
  - The accuracy on the validation data using data attribute best_score is 0.8464
  - The accuracy of test data is: 0.8333333333333334
- Support Vector Machine
  - The accuracy on the validation data using data attribute best_score is 0.8482
  - The accuracy of test data is: 0.8333333333333334
- Tree Classsifier
  - The accuracy on the validation data using data attribute best_score is 0.9018
  - The accuracy of test data is: 0.7777777777777778
- k nearest neighbors
  - The accuracy on the validation data using data attribute best_score is 0.8482
  - The accuracy of test data is: 0.8333333333333334

https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

RESULTS – MACHINE LEARNING PREDICTION

- The Accuracy of test data of Linear Regression, Support Vector Machine, and KNN are equal. The same confusion matrix is shared among these three statistical models with 3 false positive for each of them. The Tree classifier has a lower test data accuracy and 3 false negative and 1 false positive showing worse results.
- The model accurately predicted mission success



https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

CONCLUSION

An analysis of imported data from SpaceX and Wiki was performed to investigate factors driving the success of booster recovery

- Booster Version, Launch Sites, Orbit, and Payload play all a key role in the mission success
- 41.7% of all successful launches were performed at KSC LC-39A site launch
- SSO orbit was the only orbit with more than one launch and success rate of 100%
- The success rate gradually increased since 2013 until 2020 with an average success rate for the total number of launches of 0.6667
- The best models to predict mission outcomes were identified in KNN and Support Vector Machine, both presenting an accuracy of 84%

GITHUB FILES

- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb
- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/edadataviz.ipynb
- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb
- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/jupyter-labs-webscraping.ipynb
- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/lab_jupyter_launch_site_location.ipynb
- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb
- https://github.com/Lorscience/Data_Science_Capstone_SpaceX_IBM/blob/main/spacex_dash_app.py